README and Guidance

Andrew Foote, Mark Kutzbach, Lars Vilhuber

2020-09-07

This README describes the data inputs and processing stream for our paper Recalculating … : How Uncertainty in Local Labor Market Definitions Affects Empirical Findings.

Data Availability and Provenance Statements

Commuting Zone Data

Economic Research Service (2012)

- Source https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/ (https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/)
- Actual file: https://www.ers.usda.gov/webdocs/DataFiles/48457/czlma903.xls?v=6997.1 (https://www.ers.usda.gov/webdocs/DataFiles/48457/czlma903.xls?v=6997.1)
- Provided as part of this replication package

CZ data were produced by an agency of the US Government and are in the public domain.

Journey-to-Work (JTW) data

Most of the JTW data can be found at https://www.census.gov/topics/employment/commuting/guidance/flows.html (https://www.census.gov/topics/employment/commuting/guidance/flows.html). The data were produced by an agency of the US Government and are in the public domain.

Because the US Census Bureau does not provide robust (permanent) URLs, we archived the data on openICPSR/DataLumos, or searched for permanent locations elsewhere on ICPSR. As of 2020-09-01, the source URLs were still functional, though. Our scripts pull the data from the source URL.

1990 JTW

U.S. Census Bureau (2017a)

- Source: https://www2.census.gov/programs-surveys/commuting/datasets/1990/worker-flow/usresco.txt (https://www2.census.gov/programs-surveys/commuting/datasets/1990/worker-flow/usresco.txt)
- Permanent source: http://doi.org/10.3886/E100617V1 (http://doi.org/10.3886/E100617V1)
- Not provided
- Renamed to: 1990jtw_raw.txt

2000 JTW

U.S. Census Bureau (2003)

- Source: https://www.census.gov/population/www/cen2000/commuting/files/2KRESCO_US.txt (https://www.census.gov/population/www/cen2000/commuting/files/2KRESCO_US.txt)
- Permanent source: http://doi.org/10.3886/ICPSR13405.v1 (http://doi.org/10.3886/ICPSR13405.v1)
- Not provided
- Renamed to : jtw2000_raw.txt

2009-2013 ACS flows

U.S. Census Bureau (2017b)

- Source: https://www2.census.gov/programs-surveys/commuting/tables/time-series/commuting-flows/table1.xlsx (https://www2.census.gov/programs-surveys/commuting/tables/time-series/commuting-flows/table1.xlsx)
- Permanent source: http://doi.org/10.3886/E100616V1 (http://doi.org/10.3886/E100616V1)
- Renamed to: jtw2009_2013.csv
- Not provided

BEA data

Data on National Income and Product Accounts (NIPA).

Bureau of Economic Analysis (2019)

- Note: Data can be downloaded from https://apps.bea.gov/regional/downloadzip.cfm (https://apps.bea.gov/regional/downloadzip.cfm), under "Personal Income (State and Local)", select CAINC30: Economic Profile by County, then download. A direct download is also possible, see next line. The file is regularly updated.
- Source: https://apps.bea.gov/regional/zip/CAINC30.zip (https://apps.bea.gov/regional/zip/CAINC30.zip).
- Datafile: CAINC30__ALL_AREAS_1969_2018.csv
- The datafile is provided as part of this package.

The data were produced by an agency of the US Government and are in the public domain.

Dataset list

The following files are provided in `$raw` directory:

filename
CAINC30__ALL_AREAS_1969_2018.csv
jtw2009_2013.csv
popcounts.dta
table1.xlsx
urates_counties.csv

Software Requirements

- SAS 9.4 (TS1M0)
  - SAS/STAT 12.3 (maintenance)
- Stata 14.2
- R 4.0.2 (used only to automate cleaning of one data file)
  - readxl, tidyr, dplyr, readr for processing
  - rprojroot, config for configuration
  - all dependencies are installed upon first run
- Bash, Curl, wget as part of download (may require Linux, but can be replaced by manual downloading)

Memory and Runtime Requirements

These programs were last run as follows:

- OS: Linux CentOS release 6.3 (Final)
- 8-core (though probably only 1 core was in use)
- 147 GB RAM (unlikely to have been fully utilized)
- about 1.5GB disk space required

Description of programs

Setting up data

See the raw data README.md (../raw/README.md) for details in preparing the raw data files. They are not downloaded by the SAS and Stata programs in the `$programs` folder.

Main program files

filename
01_dataprep.sas
02_clusters.sas
03_prep_figures.sas
04_figures2_3.do
05_01_flows.do
05_02_bootstrap.sas
05_03_bootstrap_graphs_new.do

config.do

config.sas

statado

Setting up programs

- modify config.sas :
    - change the line with root to correspond to your project directory
- modify config.do :
    - change the line with root to correspond to your project directory

Order of programs to run

To create the replicated commuting zones, run the following programs (parameters below):

Reading in various datasets

```
sas 01_dataprep.sas
```

(runtime: 2.81s)

CLUSTERING PROCESS

```
sas 02_cluster.sas
```

(runtime: 3:25.73 minutes)

OUTPUT: $data/clusfin_jtw1990.sas7bdat

CUTOFF by CLUSTER COUNT GRAPH

```
sas 03_prep_figures.sas
stata -b do 04_figures2_3.do
```

The first program took 8:50 minutes. The second one runs in seconds.

BOOTSTRAP

Run the Bootstrap

Projects MOEs from 2009-2013 onto 1990 data, creates the 1000 realizations of commuting zones.

```
stata -b do 05_01_flows.do
sas      05_02_bootstrap.sas
```

The first program runs in seconds, the second one takes 55 hours.

Figure 4

```
stata -b do 05_03_bootstrap_graphs_new.do
```

(runtime: seconds)

Replication programs for analysis in Section 4.1

All programs are in $programs/06_qcew/ subdirectory. Change working directory.

filename
00_bea_readin.do
00_describe_bootclusters.do
00_qcew_extraction.sas
00_qcew_post_extraction.do
00_readin_czones.do
01_regressions_table.do
02_01_cluster_loop.do
02_02_cluster_loop.do

03_01_cluster_graphs.do
03_02_cutoff_graphs.do
zz_bartik_merge.do

(NOT ADJUSTED YET)

Data prep programs:

- 00.bea_readin.do
- 00.qcew_extraction.sas

Regression table:

- 01_regressions_table.do

Graphs:

- Figure 5:
  - 02.01.cluster_loop.do
  - 03.01.cluster_graphs.do
- Figure 6:
  - 02.02.cutoff_loop.do
  - 03.02.cutoff_graphs.do (creates figure 6)

Programs called during processing: zz_bartik_merge.do

Replication programs for Section 4.2

All programs in $programs/adh/ subdirectory

Data prep programs (in order):

- 00.01.IPW_creation.do
- 00.02.mergecounty.do
- 00.03.cz_merge.do
- 00.04.aggregatedata.do

Table 3 is created by:

- 01.table3.do

Figures are created by:

- Figure 7:
  - 02.02.overall_loop.do
  - 03.02.overall_graphs.do
- Figure 8
  - 02.01.cutoff_loop.do
  - 03.01.cutoff_graphs.do

List of tables and programs

(needs to read in code-check.xlsx)

References

Bureau of Economic Analysis. 2019. "Table 30: Economic Profile by County, 1969-2018." [Datafile]. U.S. Department of Commerce [producer].

Economic Research Service. 2012. "1980 and 1990 Commuting Zones and Labor Market Areas." [Dataset]. United States Department of Agriculture.

U.S. Census Bureau. 2003. "Census of Population and Housing, 2000 [United States]: County-to-County Worker Flow Files: Version 1." [Datafile]. U.S. Department of Commerce [producer].

———. 2017a. "1990 County-to-County Worker Flow Files." [Datafile]. U.S. Department of Commerce [producer].

———. 2017b. "2009-2013 5-Year American Community Survey: Commuting Flows." [Datafile]. U.S. Department of Commerce [producer].